Response to Reviewer #1's Comments

Dear reviewer,

Many thanks for your time in reviewing our manuscript and providing insightful comments and detailed suggestions. Following your constructive feedback, we revised the text and figures to strengthen the novelty and accuracy of this manuscript. Below are our detailed responses and revisions (Black: your comments; Blue: replies; *Purple Italic: revisions*).

1. This study presents a new dataset of global drainage system. The authors illustrate that the dataset has a higher resolution with acceptable accuracy and additional shape attributes. Generally, the English writing of the manuscript needs significant improvements.

Thank you for pointing out that the English writing in this manuscript needs improvement. In addition to addressing the specific language comments you raised, we have polished the English throughout the text. Furthermore, we used the Grammarly Business software to check the grammar and spelling of every sentence. Please see all revisions in the manuscript with changes tracked. Below are some examples of changes made:

(1) *Flow direction is the basis for calculating flow accumulation~~. Flow accumulation is~~, a measure ~~used in hydrological analysis to quantify~~ that quantifies the cumulative number of cells contributing to a given point.*

(2) *~~Then, we~~ We selected a basin in Madagascar to demonstrate the ~~processes~~ steps of obtaining drainage basins and their longest rivers from a DEM ~~through a series of hydrological steps~~.*

(3) *~~The river~~ River mouths can be ~~a location where the~~ river confluences ~~meets a bigger river~~, ~~a~~ lakes, or seas ~~the ocean~~.*

(4) *Strahler stream order ~~is a method used to classify and quantify~~ quantifies the hierarchy of basins and river segments within a river network.*

(5) *We report all of these eight metrics and stream order values for catchments ~~with a size larger than~~ over 50 km^2.*

There are several major concerns about the manuscript as listed below:

2. The authors did not clarify clearly about the novelty of the dataset. The processing approaches are from references and there are not any improvements from authors. There are plenty of available drainage system datasets, what are the main advantanges of Basin90?

We greatly appreciate you pointing out that the novelty of our new database was not clearly articulated in the previous manuscript. Based on your second and fifth comments, we realized there were two main reasons for not clearly articulating the novelty. First, we did not specifically point out the deficiencies of published databases. Second, we did not explain our innovations in methodology.

Therefore, **we rewrote the Introduction section** to describe the deficiencies of published databases more accurately. We clearly explained the research gap, the novelty of Basion90m, and its advantages. The last three paragraphs of the revised Introduction section are as follows (*purple* indicates new content):

The above datasets lack measurements of drainage basin length and aspect ratio. Shen et al. (2017) used 1-km DEM to obtain the global distribution of basin length and elongation ratio. However, the dataset is in raster format, without vectorial basin boundaries and rivers. More importantly, the spatial resolutions of all the above databases are relatively low. With the advancements in computer performance and algorithms, the 90-m resolution DEMs are being used to establish global databases of drainage systems. For example, the HDMA released by USGS includes nearly 295000 drainage basins but only contains information on drainage areas (Verdin, 2017). GRNWRZ comprises a global river database at a resolution of 90 m that includes information on river lengths (Yan et al., 2022). This database offers the boundaries and areas of water resource zones, which are distinct from drainage basins.

In summary, catalogs of drainage basins and rivers are available, along with measurements of area, slope, and elevation of basins. Yet, many datasets were based on DEMs with resolutions of 500 m or coarser, and few works have focused on more complex basin characteristics, such as aspect ratio, which describes the shape of drainage basins, and sinuosity that characterizes the shape of river channels. Furthermore, some databases only offer raster formats without vector accessibility. Moreover, the download links of some datasets are invalid (e.g., Vörösmarty et al., 2000; Guth, 2011).

Here, we provide an updated global catalog of drainage systems, making all relevant files and script available, and including a wide range of geometric characteristics. We used a global 90-m resolution DEM and obtained over 665000 drainage basins with a size over 50 km^2. For each basin, we extracted the longest river channel that extends from drainage divide to river mouth. Additionally, we measured parameters for each drainage system, including stream order, the length, width, aspect ratio, slope, and elevation of basins, and the length and sinuosity of rivers. The spatial distribution of drainage systems and their morphological parameters constitute a drainage system shape dataset, Basin90m.

Regarding **methodological innovation**, while we utilized published TopoToolbox functions, we developed new methods and functions to calculate basin length and width. Another innovation is the integration of all functions into an automated workflow that extracts basins and channels with multiple morphological parameters for all stream orders from a single DEM input. To better illustrate our innovation in methodology, we have added the following sentences in the Methodology Section:

(1) *In addition to utilizing the functions in TopoToolbox, we developed new algorithms to measure basin length and width. We integrated all these functions into an automated workflow that delineates drainage basins and channels across all stream orders and measures their morphological attributes.*

(2) *Basin width is defined as the greatest distance measured along a straight line perpendicular to the direction of basin length at the two points where the line intersects basin boundary.*

3. Also the importance of the dataset is not demonstrated why the shape attributes of basins are important and how they can affect runoff generation.

We appreciate your suggestion to better demonstrate the importance of the shape attributes of drainage systems. In the first paragraph of the Introduction, we stated that basin aspect ratio and river sinuosity influence landscape evolution processes and dynamics. Following your insightful comment, **we extended the discussion on how basin aspect ratio affects water level, flood risks, and the richness of species**. The revised paragraphs relating the importance of basin shape attributes are as follows (newly added in *purple*):

Drainage divides and rivers are among the most recognizable features on Earth's surface. The shape of drainage basins and rivers holds significant implications for landscape evolution processes and dynamics (Kirchner et al., 2001; Shelef, 2018; Ielpi et al., 2023). With equal basin lengths, a broader basin collects more precipitation thus offering a more stable discharge and water level, reducing the risk of river drying up. A stable water level benefits ecological integrity, water use, and navigation (Datry et al., 2023). Broader basins typically have longer and more intricate river networks, affecting flood risk associated with both extreme rainfall and glacial lake outburst events. In broader basins, rainwater takes more time to reach the main stream, indicating a longer arrival time of flood peaks. A glacial lake outburst flood travels a longer path in a broader basin, losing greater energy and surface runoff. Hence, communities downstream of a broader high-mountain catchment encounter less threat from a single glacial lake outburst flooding. However, as broader basins have more tributaries with risks of glacial lake outbursts, timely and accurate flood early warnings require deploying more seismic stations within the basin (Maurer et al., 2020; Cook et al., 2021).

Furthermore, broader basins can contain a greater variety of tributaries and habitat types. This heterogeneity of habitats allows these basins to host a greater diversity of species, especially aquatic, amphibious, and riparian organisms, by providing more ecological niches across the landscape (Matthews et al., 1998). Similarly, meandering rivers create a mosaic of habitats with varying flow velocities, depths, and substrates, supporting the diversity of aquatic organisms (Nagayama and Nakamura, 2017; Rhoads et al., 2003; Yu et al., 2022).

Meanwhile, given the same elevation drop, a meandering river has a milder channel gradient than a straight river and therefore features lower erosion rates. Additionally, the shape of drainage systems has been argued to be related to climatic and tectonic conditions (Castelltort et al., 2012; Ielpi et al., 2023; Luo et al., 2023; Sreedevi et al., 2009; Strong and Mudd, 2022), and could therefore be used as an archive to reconstruct Earth's history. Accordingly, a global dataset on the shape of drainage systems benefits scientists and policymakers in geomorphology, hydrology, and ecology, fostering interdisciplinary collaborations.

4. The validation of the dataset is far from enough only doing it in a specific basin. More validations are needed, for example comparing with measured basin area (there are many measured basin area in the US as far as I know), or a global comparison of basin area from Basin90 and HydroSheds.

According to your suggestion, we chose HydroATLAS database (Linke et al., 2019) to compare the accuracy of basin areas in Basin90m. These two databases use different methods to classify basins. HydroATLAS uses the Pfafstetter Coding System (please see figure below). Basin90m uses the Strahler stream order method. Basin90m takes the basin outlet as the starting point and includes all upstream areas in that basin. In contrast, in the Pfafstetter Coding System used in HydroATLAS, a basin can be the area between an outlet and the confluence point of its upstream tributaries. Basin #1 in figure below is such an example.

Due to these differences in basin hierarchical classification, the basins in the two databases are not always comparable. For example, basins #1, 3, 5, 7 in figure below do not have corresponding basins in Basin90m. Furthermore, if comparing globally, it is nearly impossible to systematically check whether basins in the two databases correspond, because it can just be a higher-level basin in one database containing a lower-level basin from the other database. Therefore, global comparisons are very challenging and have great uncertainty.

Figure caption: Pfafstetter coding scheme used in HydroATLAS (modified from Technical Documentation of HydroATLAS).

Therefore, we selected 10 basins across the USA for comparison (please see Fig. 9 below). To capture diverse terrain, climate, tectonics, and vegetation conditions as much as possible, these 10 basins are evenly distributed along the east-west direction of the USA. The average absolute relative error of basin areas from these 10 basins in the two databases is less than 0.5%, indicating consistency between the two databases in delineating basin boundaries.

Based on the above results, we added **a new figure** (**Fig. 9**) and **a new section** (**Section 4.3 Accuracy of basin area**). On this basis, we also made modifications to other relevant sections of the manuscript (such as the Conclusion section).

We appreciate you for providing valuable suggestions that improved our validation section. The newly added text and figure are as follows:

4.3 Accuracy of basin area

Drainage area is an important metric to evaluate the accuracy of drainage basin delineation. We compared drainage areas in Basin90m with those in HydroATLAS (Linke et al., 2019). We selected ten basins across the USA for comparison (Fig. 9a). These drainage basins are evenly distributed in the east-west direction across the North American continent, thus spanning diverse geological regimes, terrains, climates, and vegetation environments. Basins #1-5 are located in the west with arid climate, steep topography, and relatively active tectonics. Basins #6-10 are located in the east with low topographic relief but more precipitation and vegetation.

Due to the difference in DEM resolution (90 m for Basin90m and 500 m for HydroATLAS) and algorithms for delineating basin boundaries, the drainage divides from the two databases *do not always coincide (Fig. 9b). But the two databases are consistent without substantial discrepancies (Fig. 9c). We quantified their difference using absolute relative error. The absolute relative error is the absolute value of the area difference between Basin90m and HydroATLAS as a ratio to HydroATLAS. Smaller values indicate higher agreement between the two datasets. Except for a 1.2% absolute relative error for basin #4, the values for the other nine basins are below 0.8%. The average absolute relative error for the ten basins is 0.47% (Fig. 9c). This slight discrepancy is acceptable given the nearly five-fold difference in DEM resolution and variations in algorithms for basin delineation.*

In summary, the validation results based on the Peruvian Moche River basin and ten representative basins spanning east-west across the USA indicate that Basin90m has a high resolution for basin boundaries and river channels. Besides, the morphological parameters exhibit high accuracy.

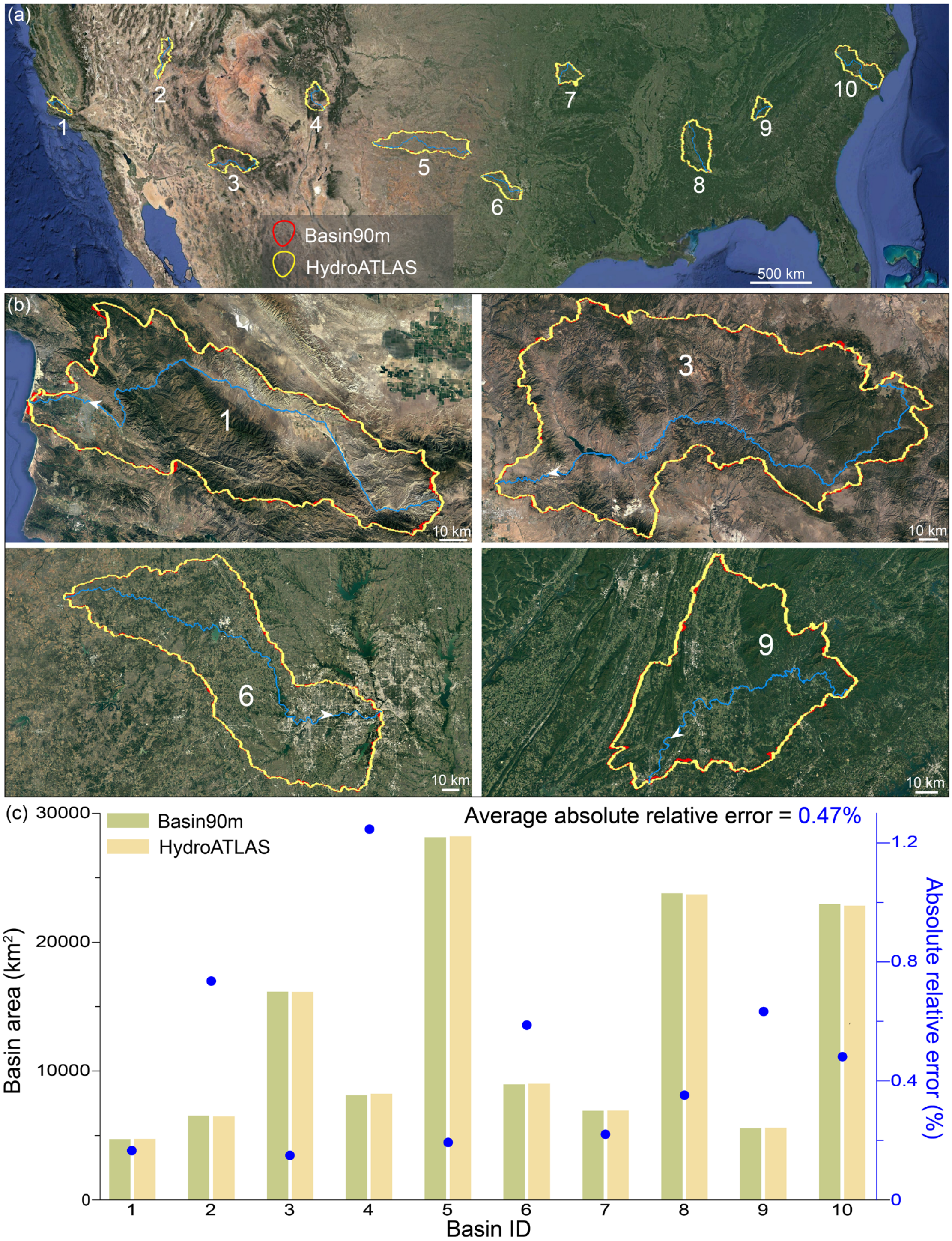

Figure 9. *The accuracy of basin area in Basin90m, compared against HydroATLAS (Linke et al., 2019). The stream order for all the ten representative drainage basins is four. (a) Google Earth image shows ten example drainage basins in the USA. (b) Enlarged images of four representative basins. Basin #1 has an ocean outlet. Basin #3 is situated in a high-altitude arid region. Basin #6 features a flat terrain and encompasses a major city (Dallas). Basin #9 is located in a tectonically active folded region. (c) Comparison of basin areas between Basin90m and HydroATLAS.*

There are also some other comments as below:

5. Introduction is weak that authors did not present sufficient evidence of the shortcoming of previous datasets.

Thank you for suggesting that we should directly point out the deficiencies of published databases, in order to better demonstrate the novelty of Basin90m. We inserted the following sentences in the Introduction section to point out the shortcomings of previous datasets:

(1) *The above datasets lack measurements of drainage basin length and aspect ratio.*

(2) *However, the dataset is in raster format, without vectorial basin boundaries and rivers.*

(3) *More importantly, the spatial resolutions of all the above databases are relatively low.*

(4) *For example, the HDMA released by USGS includes nearly 295000 drainage basins* ***but only contains information on drainage areas*** *(Verdin, 2017).*

(5) *Yet, many datasets were based on DEMs with resolutions of 500 m or coarser.*

(6) *Furthermore, some databases only offer raster formats without vector accessibility.*

(7) *Moreover, the download links of some datasets are invalid (e.g., Vörösmarty et al., 2000; Guth, 2011).*

6. There are already some datasets that include drainage basins from 90m DEM, why did authors start from the raw DEM datasets rather than using the existing drainage basin datasets to extract the shape of the drainage basins and sinuosity of rivers?

Our main goal is to provide multiple shape parameters for global basins and river channels, as well as the topological structure, namely stream order. Existing global databases that based on 90-m DEMs do not simultaneously meet the above conditions.

For example, the HDMA (Hydrologic Derivatives for Modeling and Analysis, Verdin, 2017) released by the USGS contains about 295000 drainage basins. However, the river channels in this database do not start from drainage divide, but begin where the upstream area is greater than 250 km^2. This means that we cannot use the river channel in HDMA to calculate basin length and width, nor can we calculate river length and sinuosity.

Therefore, by constructing a Matlab code, we fully automatically extract river channels (from drainage divide to outlet) and drainage basins, and calculate various parameters and stream order. In addition, by choosing a smaller drainage area threshold (50 km^2), we obtained 670000 basins and rivers, far more than the quantity in published datasets.

7. Line 178, I don't think basins with an area <50km2 do not contribute to the downstream basins if they are connected, like the headwaters. Please clarify what authors mean here.

The commented sentence is:

In Basin90m, basins with an area of less than 50 km^2 were not extracted, thus not contributing to the stream order of its downstream basins.

We intended to convey that small basins with an area of less than 50 km^2 have not been extracted. Therefore, in Basin90m, these small basins do not contribute to an increase in the stream order of downstream rivers.

We agree with you that all rivers, regardless of their size, contribute to downstream water discharge. To clarify this, we have added the following sentence:

However, those small basins will contribute to the water discharge of their downstream channels.

8. Line 450, What size is the Moche River basin for comparison? Please provide more information about the basin.

We are grateful to you for suggesting that more basic information about the Moche River basin should be included. Accordingly, we have inserted the following sentence in the first paragraph of Section 4.1:

Moche River basin encompasses an area of 2143 km^2, with its highest altitude reaching 4257 m.

9. There are many typos or mistakes. Line 23, “Google”, not ‘Goole’. Line 121 ‘to obtain’. Figure 2(c) ‘river mouth’. Please check through the manuscript.

We sincerely appreciate this valuable feedback. We have made the corrections.

In Line 23, we have corrected “Goole” to “Google”.

In Line 121, we have replaced “to obtaining” with “to obtain”.

In Figure 2c, we have revised “river month” with “river mouth”.

In addition to these specific changes, we have conducted a thorough language review of the entire manuscript, including both text and figures, to ensure quality and accuracy.